DOCUMENT RESUME

ED 322 630 EA 022 155

AUTHOR Purser, Susan R.; And Others
TITLE The Relationship between Secondary Principals' Evaluation of Teachers and Pupil Achievement.
PUB DATE Apr 90
NOTE 32p.; Paper presented at the Annual Meeting of the American Educational Research Association (Boston, MA, April 16-20, 1990).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Academic Achievement; Correlation; High Schools; Job Performance; Measurement Techniques; Predictor Variables; *Principals; Regression (Statistics); Statistical Inference; *Teacher Administrator Relationship; Teacher Characteristics; *Teacher Effectiveness; *Teacher Evaluation; Teacher Supervision

ABSTRACT

The relationship between teacher effectiveness, as measured by residualized student grades, and teacher evaluations by secondary principals is analyzed in this study. The two-step methodology involved baseline data collection and an effectiveness development phase. In the first phase, demographic variables were obtained from personnel files of 145 teachers and from school district students enrolled in English, mathematics, science, and social studies. Data were also collected from principals' teacher evaluation reports. The effectiveness development phase established regression coefficients for student demographic variables and classified levels of teacher effectiveness from a stratified random sample of 87 high school teachers. Multiple discriminant analysis was then used to test the relationship between principals' evaluations of teachers and teacher effectiveness based on student achievement after controlling for certain student variables. Findings indicate that the use of a secondary principals' evaluation process is teachers (44 black, 43 white; 54 female, 33 male; content areas of English, mathematics, science and social studies). Multiple discriminant analysis was then used to test average, average, and below average. However, principals' evaluations effectively identified the most effective teachers. Recommendations are made for: (1) inclusion of student contextual factors in teacher evaluations; (2) assessment of pupil achievement prior to conducting teacher training and evaluation; and (3) application of the model to evaluations of buildings and systems and to elementary-secondary education. Statistical tables are presented. (12 references) (LMI)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED322630

THE RELATIONSHIP BETWEEN SECONDARY PRINCIPALS' EVALUATION OF TEACHERS AND PUPIL ACHIEVEMENT

Susan R. Purser, Ed. D.
Director of Planning, Research and Information
Services Division
Oklahoma City Public Schools

Harold V. Knight, Ed. D.
Professor of Educational Leadership and Research and
Director of Research Training
University of Southern Mississippi

Edgar H. Bedenbaugh, Ed. D.
Chair of Development of Educational Leadership and Research
University of Southern Mississippi

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Susan R. Purser

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper prepared for presentation at
American Educational Research Association
Annual Meeting, Division K
Boston, Massachusetts

April 16-20, 1990

EA 022 155

BEST COPY AVAILABLE

THE RELATIONSHIP BETWEEN SECONDARY PRINCIPALS' EVALUATION OF TEACHERS AND PUPIL ACHIEVEMENT

Susan R. Purser, Ed. D.
Director of Planning, Research and
Information Services Division
Oklahoma City Public Schools

Harold V. Knight, Ed. D.
Professor of Educational Leadership and Research and
Director of Research Training
University of Southern Mississippi

Edgar H. Bedenbaugh, Ed. D.
Chair of Development of Educational Leadership and Research
University of Southern Mississippi

Teacher effectiveness is a call word of the times. Teacher evaluation is a process that has been going on informally for generations. How do these two terms "fit" together? How should teacher effectiveness be evaluated? Have methods of today and yesterday done a bad job in determining who has an impact on students? Does effectiveness translate into student outcomes -- student achievement?

The term teacher effectiveness is being used quite frequently by the researchers as well as the public. Each use, however, seems to carry with it a different meaning. Even though the attempts at determining the distinguishing characteristics have not been completely successful, there is no reason to reach the conclusion that the questions of teacher effectiveness cannot be answered (Doyle, 1978).

Research in the area of teacher effectiveness has moved from observation and description to correlational studies to experimentation. Teaching practices were observed and recorded.

Later those practices were correlated to student achievement. Now those same practices are being manipulated as determinations are being made about their impact on student achievement (Gage, 1984).

In 1978, school administrators indicated that from 5 to 15 percent of the teachers within their districts were not performing up to an acceptable level (Gudridge, 1980).

Edwin M. Bridges, in an editorial for Education Week, November 20, 1985, very pointedly summarized the impact:

> Incompetent teachers represent a relatively small proportion of the work force, but the number of students who are being taught by such teachers is substantial. Even if only 5 percent of the teachers in public elementary and secondary schools are incompetent, the number of students being taught by these teachers exceeds the combined public school enrollments of the 14 smallest states.

As a promise of hope for the field of education, one of the implications of research is that most ineffective teachers can be helped to improve when a proper evaluation tool is used (Rosenholtz, 1984).

The primary purpose of this study was to ascertain the degree to which secondary school principals' evaluations of teachers relate to the outcome variable of their pupils' achievement. Although researchers have identified teacher behaviors that are purported to be indicators of teacher effectiveness, there is a paucity of research investigating the relationship between these descriptors and student achievement (Rosenshine and Furst, p. 66). Such research studies should benefit the educational process but are very difficult to replicate in the actual field setting.

The increased emphasis on teacher effectiveness is resulting from policy statements which are typically drafted due to the public outcry for the improvement of teaching in the public and private schools. The emphasis has shifted to consider how effective teaching takes place. Lists of indicators of teacher effectiveness have been developed and utilized in controlled situations (Doyle, 1978). However, unless reputable findings which link teacher behaviors to student achievement in ordinary classrooms is found, the evaluations of teachers will not be believable (Berliner, 1976).

Therefore, in an effort to advance this knowledge base, this paper represents the results of a study designed to include contextual factors in measuring teacher effectiveness based upon pupil cognitive outcomes. The relationship between these "adjusted" measures and the evaluation ratings of secondary teachers by their principals was examined. The evaluation cycle was a comprehensive process which was based on behaviors of effective teaching. A study by Purser (1986), utilizing the same subjects as the current study, found that no significant relationships existed among teacher demographic variables, student characteristics, and evaluation ratings of secondary principals. These findings provided a baseline for this study. This follow up analysis provided for a more concentrated emphasis on the of the degree to which teacher effectiveness (based on pupil outcomes) could be predicted by the use of a comprehensive summative evaluation system by secondary school principals.

The ultimate goal of the study was to provide data which may assist school districts in improving the process of predicting and assessing teacher effectiveness. The study was conducted using subjects from an urban school district with a student population of approximately 30,000. The district was approximately 70% minority.

This study used regression techniques to develop "expected" pupil outcome measures and then to identify levels of teacher effectiveness by using residualized measures (actual v. expected achievement means). The ability of the principals' evaluation ratings to discriminate between three established levels of teacher effectiveness was assessed. This detailed data analysis technique was utilized in order to include more accurately available contextual variables. The authors present evidence that pupil outcome measures can and should be one component of a model for teacher evaluation.

Method

The methodology consisted of two distinct phases necessary to test the major hypotheses of this study. These two phases consisted of (1) baseline data collection of information on pupils, teachers, and principals' evaluation results to assess the levels of teacher effectiveness and, (2) effectiveness development that used the baseline data in defining teacher effectiveness levels. The final analysis of data consisted of using the secondary principals' summative evaluation data to determine the degree to which the evaluation process data could accurately classify

teachers into teacher effectiveness levels based on "adjusted" pupil achievement scores.

Baseline Data

Baseline data for this study was from a population of 145 teachers of English, mathematics, science, and social studies in grades 10-12 of the school district. Each teacher had a valid teaching certificate awarded by the state department of education. The sample included 74 minority teachers and 71 non-minority teachers, with 53 of the teachers being males and 92 females. The following represents the data collection protocol used in this phase of the study.

1. Teacher data consisting of race, gender, level of certification, area of certification, years of teaching experience, and specific content area course(s) taught were obtained from personnel files.
2. Student files were used to obtain the student variables consisting of race, gender, socioeconomic status, scores on district standardized subject area tests for courses in each of the four content areas, and final grades in each of the courses.
3. Data representing the principals' evaluation of teachers as reported through the district's summative evaluation report was collected from personnel files. These data included ratings

on teachers in each of the four performance areas of instruction, classroom management, personal relations, and professional responsibilities and a composite rating consisting of the sum of the four performance area ratings.

Effectiveness Development

This phase involved extensive data analysis techniques consisting of establishing regression coefficients for the student demographic variables to use in predicting pupil achievement. These weights were then used in a cross validation procedure with a random sample of pupils in each course within each content level. This procedure yielded a more stable regression weight for each student variable in predicting grades in each course taught in the four content areas. The predicted grades for the students represented the districtwide "expected performance level" for pupils taught by the teachers included in this study. These districtwide performance levels were then used to determine "expected" pupil achievement (grades) enrolled in classes of the teachers selected for the final phase of the study.

In order to establish teacher effectiveness levels, the following analysis steps were conducted: (1) the "expected" pupil achievement was determined for each class taught by the final 87 teacher subjects; (2) a confidence interval was established around the total predicted mean that represented .675 standard error units above and below the estimated mean (this unit was used to

categorize the final effectiveness levels so that approximately 25% of the teachers would be placed at the highest and lowest levels respectively); (3) the actual obtained mean for each course taught by each teacher was calculated and compared with the "expected" mean resulting in a residualized measure that was used as the final criterion variable; and (4) the classification of the teachers into three effectiveness levels representing the degree to which actual pupil performance deviated from "expected" pupil performance for each course after adjustments were made for pupil and contextual content variables.

This extensive process provided a basis for establishing three teacher effectiveness levels: (1) Below Average -- consisting of actual means falling .675 standard error units below the predicted mean, (2) Average -- consisting of actual means falling within the range of .675 standard error units above and below the predicted mean, and (3) Above Average -- consisting of teachers with actual class means exceeding .675 standard error units above the predicted course.

Multiple discriminate analysis was used as the primary data analysis technique by using the three (3) established teacher effectiveness levels as the criterion variable with the principals' evaluation scores on each of the performance areas serving as the predictor variables. This technique provided a basis for testing the relationship between principals' evaluation of teachers and teacher effectiveness based on the achievement (outcome) of pupils taught by each teacher after adjusting for selected pupil

variables. Stepwise discriminant analysis was used with additional one-way anovas run using the discriminant scores of the teachers as the dependent variable and the effectiveness levels as the grouping variable. Such anovas are often of assistance to consumers of studies of this nature who may not have been recently exposed to the terminology of discriminant analysis. Finally, total and within content area analyses were conducted to provide for differential analysis by content area. The .05 alpha level was used as the rejection level in all inferential tests.

Subjects

The subjects consisted of 87 secondary level school teachers in the content areas of English, mathematics, science, and social studies, in the school district. A stratified random sample technique was used in the selection of these teachers using the variables of specific classes being taught within teachers discipline, certification area, race, and gender. Table 1 presents the results of this sampling procedure. These data show that there were a total of 44 black and 43 white teachers with 54 being female and 33 male. Inspection of these data according to certification area show that approximately an equal number of teachers in mathematics and science were selected according to race and gender. The certification area of English showed that the teachers were predominately black and overwhelmingly female (18 females, 2 male). The social studies teachers were predominately white and female as evidenced by the selection of 14 white and 14 female teachers for this stratification procedure.

Insert Table 1

Table 2 displays the distribution of the teacher sample according to the certification area, level of certification, and effectiveness level. Approximately equal numbers of teachers were selected according to level of certification and certification area in the three effectiveness levels. Therefore, these data indicate that the use of the confidence interval method for establishing the teacher effectiveness levels was adequate. The descriptive data representing the teacher sample presented in Tables 1 and 2 verify that this sample was representative of the total district's secondary school level characteristics. In other words, all subjects selected were proportional to the actual number of secondary teachers within each content area for the entire school district.

Insert Table 2

Results

Descriptive data representing the means of standard deviations of the principals' evaluation scores for the teachers in each of the performance areas of instruction, classroom management, interpersonal relations, professional responsibilities, and the

total evaluation score are presented in Table 3. Inspection of these data show that the highest total evaluation mean was for the teachers classified as below average, with the lowest total mean being for the above average level. This trend was also evidenced in the interpersonal relations and professional responsibilities categories. This pattern did not exist for the instruction and classroom management performance areas. The variability of the evaluation scores was very small as the highest standard deviation was .39 for the average effectiveness level in the interpersonal relations performance area and lowest (.19) in the performance area of instruction for the below average effectiveness level. It may be further observed that for the two evaluation performance areas of instruction and classroom management, the variability was very low in comparison with the other performance areas. The data in Table 3 show a standard deviation of .32 being the highest for the instruction and classroom management performance areas in the average effectiveness level. The lack of variability in the instruction and classroom management performance areas may indicate a more "generalized" rating on the part of the principals possibly because of the minimal observation time available during the school year. Such an observation must be verified by the data analysis as such descriptive data can often be misleading particularly when measuring multivariate dimensions.

Insert Table 3

The technique of multiple discriminant analysis was used to determine the power of the principals' rating in each evaluation performance area to classify teachers' effectiveness among the three levels of below average, average, and above average. Table 4 presents the results of the discriminant analysis for the total sample and by content area. These data indicate that only for the total group with a Chi square = 13.49 and $p = .036$ and for the functions for the social studies group (Chi square = 12.01, $p = .006$) were the evaluation scores significant at the predetermined .05 alpha level.

The most important information presented in Table 4 is the classification results. Since statistical significance may be influenced by sample size, the ability of the evaluation performance ratings to classify teachers correctly into the teacher effectiveness levels is of primary importance. The data in Table 4 show that for the total group a "flip of a coin" would probably classify the total group into the effectiveness categories as well as the statistical values (49.43% classified correctly). The only content area that showed a greater percent of correct classifications was the social studies group with a correct classification of 66.67% of the teachers. Obviously these data give little credence to the ability of secondary principals within this setting to utilize this evaluation process to identify teachers whose students were achieving at their expected achievement levels.

Insert Table 4

Additional Analyses

Because of the lack of discriminant power of the principals' evaluation in any of the four performance levels for the four content areas, an additional analysis consisting of combining the below average and average teacher effectiveness categories and comparing this combined category with the high effectiveness level, was undertaken to determine if the principals' evaluation process could at least discriminate between the collapsed effectiveness levels. In other words, could the most effective teachers be identified through this evaluation process?

The discriminant analysis results indicating the power of the identification using the combined levels (below and average) with the high effectiveness level group is presented in Table 5. The results of this analysis indicate that when the high effectiveness group was compared with all other groups, the four evaluation performance areas as rated by the principals were much more powerful predictors of the most effective teachers. This may be confirmed from the data in Table 5 as significance was found for each of the content areas of English and social studies and as the results for the math and science groups approached significance (.067 and .054, respectively). Likewise, the data indicating the discriminant power of the evaluation performance areas for the

total group was significant at the .05 level (Wilks Lambda .9343, Chi square 3.81, p .047).

Insert Table 5

The data for the percent of teachers classified correctly were of some note. The table with separate content area classification table shows that 72.3% of the science teachers were classified correctly, 76.17% of the social studies teachers were correctly classified, 76.19% of the math teachers, and 95% of the English teachers were correctly classified. In addition, the percent correct for the total group increased from below 50% when three effectiveness levels were used to a significant 65.52% when the high effectiveness level was compared with the combined low and average levels.

Table 6 presents an additional discriminant analysis using the average and high levels combined. These data indicate that this combination of average and high groups vs. the low teacher effectiveness level did not produce correct classifications using the four evaluation performance areas ratings by the principals. As may be observed, these data indicate that no variables met the .01 tolerance level specified for the content areas of English and math and therefore no percentage of correct classification existed. In addition, the remaining performance levels of science and social studies, while having a relatively high percentage of teachers correctly classified (77.23% and 70.83%, respectively), neither

were significant at the .05 level. Finally, when the total group of teachers was analyzed, only 58.62% of the teachers were correctly classified in the high or average combined group or the low effectiveness level.

Insert Table 6

These results yielded different conclusions from the previous findings when the low and average levels were combined into one group. The inability to identify the low effectiveness level when the remaining two levels were combined may indicate the inability of the evaluation process to identify teachers that are not performing at the average classification level. This lack of ability to identify the below average teachers does not provide very strong evidence for the time and money used in the evaluation process.

Entry Order of Variables

Table 7 presents the stepwise entry of the performance areas in order of importance in classifying teachers into the combined low and average group and the high effectiveness level. Inspection of these data in Table 7 show that for four of the five analyses performed, the evaluation performance area of classroom management was the first variable selected, indicating that this variable was perceived as extremely important by the principals. The only area in which classroom management was not selected as the first variable was in the content area of math where professional

responsibilities was the first variable selected. It should also be observed that the variable of instruction was not selected as the second variable in English and science. The second variable selected for Math was personal relations and in Social Studies the total evaluation score was the second variable selected in the step wise procedure.

Insert Table 7

Discussion

This study addressed the relationship between teacher effectiveness and the results of secondary principals' evaluation of teachers. The variable of teacher effectiveness consisted of developing three levels of effectiveness based on residualized measures representing the difference between expected and actual grades.

The first question addressed by this study was whether the use of a secondary principals' teacher evaluation process could effectively classify teachers in the three teacher effectiveness levels of above average, average, and below average. Multiple discriminant analysis was used to provide data to answer this question. The dependent variable was the three teacher effectiveness levels and the independent variables were the principals' teacher evaluation scores in each of the performance areas of instruction, classroom management, personal relations, professional responsibilities, and the composite of these four

areas. The results of this analysis indicated that the evaluation scores did significantly differentiate between the teacher effectiveness (p = .036). In addition, significance was found in the content area of social science (p = .001). However, although statistical significance was found, the most important effect, measured by the percent of teachers correctly classified, was low as evidenced by the 49.43% correct classification for the total group and 66.67% for the highest percent for a content area (social science).

The lack of classification power of the principals' rating may have been because of the very low variability in the performance areas, particularly the instruction and classroom management areas that had means in the a very restricted range. However, regardless of the statistical explanations, a more valid reason for these results may be the fact that the evaluation process was not useful in identifying effective teachers as measured by the cognitive outcomes of their pupils.

The second question was whether the principals' evaluation results could effectively identify the <u>most</u> effective teachers (above average) as compared to the remaining teachers consisting of the combined average and below average effectiveness levels.

The discriminant analysis results indicated that the principals' evaluation ratings were much more accurate in the classifying of teachers in these two effectiveness groups (total group, p = .047; English, p = .007; and social science, p = .019). Also, of prime importance, the percent of teachers correctly

classified was much higher than the results utilizing the three separate effectiveness levels (percent classified correctly ranged from a low of 79.1% for social science teachers to a high of 95% for English teachers).

These results strongly confirm the premise that teacher evaluation processes should be content specific and include learner characteristics and other measurable contextual factors. These findings are supported by the concept of teacher evaluation by the Teacher Assessment Project. Shulman (1987) in describing the basic tenants of the project not only criticizes the heavy trend toward the use of only teacher behaviors in the evaluation process but also emphasizes the need for considering such factors as (1) content specific variables, (2) characteristics of learners, (3) the contextual setting, and (4) the ability of teachers to justify specific actions-decisions in any plan for assessing teacher effectiveness.

The third question investigated in this study was if the results indicate that the principals' evaluation process correctly classified teacher effectiveness, which evaluation performance areas contributed to the discriminant power. The only positive results were obtained from the earlier described two level analysis. Therefore, the order of variable contributions to the discriminant power were presented for this two group analysis.

The evaluation performance area of classroom management was the first variable (most powerful) selected in four of the five analyses. Only for the content area of mathematics, a different

performance area (professional responsibilities) was the first variable selected. Instruction was the second performance area selected in the disciplines of English and science while the second area selected in mathematics and social science was personal relations and total evaluation. These results seem to indicate that the principals perceived management as extremely important in evaluating teachers.

Although, not quantifiable within the context of this study, the power of this management performance area possibly was so rated as a result of the teachers exercising classroom control -- thus reducing the amount of time and effort that the principals had to exert in assisting teachers in problems related to classroom management. Such a rationale may not be found in the textbooks related to evaluation, but it is certainly plausible in the "real-world" settings in which principals and teachers must function. In addition, it is plausible to project that the teachers perceived that classroom management was the primary element in the evaluation of their teaching.

Although teachers may not have been directly informed that management was the primary variable, the data do suggest that there was a congruence between the teachers and principals relative to the importance of this factor. Such a conclusion is supported by Medley (1979) in emphasizing the fact that if teachers know the criteria on which decisions affecting their careers are based, they will meet the criteria if humanly possible.

Additional results of this study indicated that the evaluation ratings of the principals could not accurately discriminate between teachers with pupils achieving at a degree far less than expected when compared with similar pupils enrolled in the same classes of other teachers. This finding was disturbing as any teacher evaluation process should identify those teachers most in need of assistance. Such results suggest that possibly the trend in evaluation should be reversed: (1) determine (by using appropriate methods including contextual factors) the relative achievement of pupils of the teacher, (2) observe the teacher in the classroom using other methods such as the portfolio scheme, and (3) mutually design a plan for improving the effectiveness of the individual teacher with clearly defined methods and criteria.

Summary

The rationale for this study of the relationship between teacher effectiveness, as measured by "residualized" pupil grades, and the teacher evaluation ratings of secondary principals may be illustrated by the remarks of Berliner (1985): "An effective teacher is one who delivers, either empirically or logically, a curriculum that is linked to outcomes."

Results of this study showed that secondary principals' ratings of teachers in the evaluation performance areas of instruction, classroom management, personal relations, and professional responsibilities did not have the discriminant power to classify teachers accurately in three levels of teacher effectiveness determined by residualized measures representing the

difference between actual and expected pupil achievement (grades). However, when the below average and average teacher effectiveness levels were combined and used in a two group analysis with the above average effectiveness level, the predictive power of the principals' evaluation ratings improved as the percent of teachers accurately classified in these two groups increased significantly. At first glance these findings may be very impressive but the results are tempered by the fact that, although the statistics support the results, the old axiom that "anyone knows a excellent teacher when he or she sees one" may be appropriately applied to these findings. The results of an additional two group analysis consisting of the combination of the levels of average and above average and the below average effectiveness level indicated that teachers classified as below average could not accurately be identified based on the ratings of the principals.

The most positive results of this study included the clear findings that effective teachers could be identified at a significantly higher level when evaluated according to specific academic content areas. Also, the methodology used to determine the "expected" pupil achievement levels using available learner characteristics and content level contextual factors proved to be appropriate as a model for including these and other factors in determining pupil outcomes. When properly utilized and interpreted, the authors support the use of this technique in differential settings and the inclusion of any appropriate predictor variables in determining expected pupil-school outcomes.

This methodology is supported by researchers such as Medley (1979) who emphasized the need for controlling for relevant student characteristics in measuring outcomes as the basis for teacher effectiveness.

The results of this study provide a basis for the following recommendations for further research in this area: (1) replication at the elementary and middle school level in order to determine contextual factors that are unique to students, teachers, and administrators at those levels; (2) the utilization of additional learner variables such as immediate and extended family variables; (3) the expansion of the outcome measures to multiple measures by including other achievement and affective domain variables; and (4) the application of this "residualized" technique to measures of building and/or system effectiveness that would facilitate the establishment of multivariable goals for such units cooperatively planned by parents, pupils, teachers and administrators. Such use of this methodology would provide for measures reflecting the total organization effectiveness which is emphasized by administration theorists but rarely empirically conducted in the "real-world" setting.

A final recommendation is that strong consideration be given to the determination of the degree to which individual teachers' pupils are meeting "expected" multiple outcomes prior to spending a vast amount of time and funds on the necessary training and conduction of detailed observational evaluations. It is recognized that this reverses the current mode of evaluation processes but the

results of this study and other research do not provide a reliable basis for the use of "indicators" of teacher effectiveness in making crucial personnel decisions including the type and nature of staff development designed to improve teaching. Such findings serve as a basis for encouraging educators to examine carefully the value of current teacher evaluation outcome data relative to its objectivity and specific utilization in the improvement of the educational process at local/state levels. This recommendation is supported by the recent work of Hanushek (1989) as the result of his research, while primarily investigating the effect of differential expenditures on pupil achievement, provided a basis for his stating: "Moreover, the dramatic differences in teachers' performance have not been captured by any account of differences in their background or classroom behaviors."

TABLE 1

Descriptive Data Representing the Distribution of Teacher Subjects on the Variables of Race, Gender and Certification Area

Certification Area	Race		Gender	
	Black	White	Female	Male
English	12	8	18	2
Mathematics	11	10	11	10
Science	11	11	11	11
Social Studies	10	14	14	10
Total	44	43	54	33

N = 87

TABLE 2

Distribution of Teacher Sample According to Area of Certification, Level of Certification by Calculated Effectiveness Category

Certification	Below Average	Average	Above Average
Area			
English	5	10	5
Math	5	11	5
Science	5	13	4
Social Studies	5	13	6
Total	20	47	20
Level of Certification			
A	7	21	5
AA	10	21	10
AAA	3	5	5
Total	20	47	20

TABLE 3

Mean and Standard Deviation of Principals' Rating of Teachers According to the Three Effectiveness Levels

Evaluation Category	Effectiveness Categories		
	Below Average	Average	Above Average
Instruction			
Mean	3.19	3.21	3.21
SD	.19	.22	.23
Class Management			
Mean	3.12	3.26	3.12
SD	.20	.32	.27
Interpersonal Relations			
Mean	3.44	3.37	3.40
SD	.38	.39	.35
Professional Responsibilities			
Mean	3.42	3.36	3.21
SD	.31	.35	.29
Total			
Mean	48.78	48.72	48.17
SD	2.27	3.48	3.10

TABLE 4

Results of Discriminant Analysis of Principals' Rating in Each Evaluation Performance Area in Identification of Teacher Effectiveness for Total Sample and by Content Area

Function	Eigenvalue	Percent of		Wilks	Chi	p
Total Group						
0						
1	.1239	69.9	.332	.854	13.49	.036*
2	.0583	100.0	.225	.949		
Classified Correctly = 49.4%						
English						
0	No Function Derived					
1						
2						
Math						
0				.7340	5.41	.267
1	.3579	99.10	.5134	.9948	.052	.815
2	.0033	100.0	.0570			
Classified Correctly = 52.36%						
Science						
0				.7340	5.01	2.47
1	.381	98.9	.5072	.9968	.06	.82
2	.003	100.0	.0670			
Classified Correctly = 52.36%						
Social Studies						
0				.4063	18.01	.006*
1	1.09	85.95	.7221			
2	.79	100.00	.3887			
Classified Correctly = 66.67%						

*Significant at .05 Alpha Level

TABLE 5

Results of Discriminant Analysis of Evaluation Categories Power to Identify Teacher Effectiveness for Combined Low and Average Group and High Expected Outcome Group

Function	Eigenvalue	Percent of Variance	Correlation	Lambda	Square	
Total Group						
0				.9343	.81	.043*
1	.0479		.2137	.854	13.49	.036*
2	.0583	100.0	.225	.949		
Classified Correctly = 65.52%						
English						
0				.3290	15.56	.004*
1	2.0396	100.0	.8192			
Classified Correctly = 95%						
Math						
0				.7401	5.419	.067
1	.3513	100.0	.5098			
Classified Correctly = 76.19%						
Science						
0				.7360	5.83	.054
1	.359	100.0	.514			
Percent Classified correct = 72.3%						
Social Studies						
0				.6873	7.89	.019*
1	.4551	100.0	.5593			
Classified Correctly = 79.17%						

*Significant at .05 Alpha Level

TABLE 6

Discriminant Analysis REsults of Using Principals' Evaluation in Classifying Subjects in the Combined Average and High Group and the Low Effectiveness Group

Function	Eigenvalue	Percent of Variance	Correlation	Wilks Lambda	Chi Square	p
English						
0	* No Functions					
1						
2						
Math						
0	* No Functions Meet Tolerance Level					
1						
2						
Science						
0				.7601	5.075	.156
1	.3156	100.0	.4898			
2						
Classified Correctly = 77.23%						
Social Studies						
0				.7635	5.609	.059*
1	.3098	100.0	.4864			
2						
Classified Correctly = 70.83%						
Total Group						
0				.9463	4.71	.107
1	.0507	100.0	.2317			
Classified Correctly = 58.62%						

TABLE 7

Stepwise Entry of Performance Areas When Teacher Effectiveness Was Categorized as Combined Low and Average and High By Content Area

	Variable Selection Order		Wilks Lambda	p
English	1.	Classroom Management	.3293	.004
	2.	Instruction	.9128	.252
Math	1.	Professional Responsibilities	.7401	.066
	2.	Personal Relations	.8353	.068
Science	1.	Classroom Management	.7365	.054
	2.	Instruction	.8402	.065
Social Science	1.	Classroom Management	.687	.011
	2.	Total Evaluation	.763	.016
Total	1.	Classroom Management	.9543	.051

REFERENCES

Berliner, D. C. (1976). A status report on the study of teacher effectiveness. Journal of Research in Science Teaching, pp. 360-382.

Berliner, D. C. (1985). Teacher effects and engaged time. Seminar conducted for Jackson Public School District, Jackson, MS.

Bridges, E. M. (1985). Coping with incompetent teachers. Education Week, V(12), 24.

Doyle, W. (1978). Interpreting teaching effectiveness research. Viewpoints in Teaching and Learning, 54(4), 141-153.

Gage, N. L. (1984). What do we know about teaching effectiveness? Phi Delta Kappan, 66(2), 87-93.

Gudridge, B. M. (1980). Overview. In Teacher competency: Problems and solutions, AASA Critical Issues Report (pp. 5-8).

Hanushek, E. A. (1989). The impact of differential expenditures on school performance. Educational Researcher, 18(4), 45-51.

Medley, D. (1979). The effectiveness of teachers. In P. Peterson & H. J. Walberg (Eds.). Research on teaching (p. 16). Berkeley, California: McCutchan Publishing Co.

Purser, S. J. R. (1986). The relationship between teacher effectiveness and teacher evaluation and selected teacher demographic variables. (Doctoral dissertation, University of Southern Mississippi, 1986).

Rosenholtz, S. J. (1984). Conclusions. MYTHS: Political myths about reforming teaching, a paper prepared for the Education Commission of the States, (No. TQ84-4), p. 22.

Rosenshine, B. V. & Furst, N. (1976). Research on teacher performance criteria. In B. O. Smith (Ed.), Research in teacher education: A symposium, (p. 37). Englewood Cliffs, New Jersey: Educational Technology Publications.

Shulman, L. S. (1987). Assessment for teaching: An initiative for the profession. Phi Delta Kappan, 69(1), 38-44.